The Stardinates - Visualizing Highly Structured Data

Monika Lanzenberger and Silvia Miksch
Vienna University of Technology
Institute of Software Technology and
Interactive Systems
Favoritenstr. 9-11/188
1040 Vienna, Austria
{monika, silvia}@asgaard.tuwien.ac.at

Margit Pohl
Vienna University of Technology
Institute for Design and
Assessment of Technology
Favoritenstr. 9-11/187
1040 Vienna, Austria
margit@igw.tuwien.ac.at

Abstract

The Stardinates are a novel interactive Information Visualization (InfoVis) technique which aims at visualizing highly structured data. They represent some Gestalt principles very well, especially the principles of Closure and 'Prägnanz'. As a consequence, Stardinates form very distinct and memorable patterns which make abstraction and aggregation much easier. We give a formal description of the Stardinates as a basis for implementation. Furthermore, we show an application by visualizing psychotherapeutic data derived from a clinical study on anorectic girls.

1 Introduction

Various studies or simulations provide a huge amount of complex data, which need to be analyzed according to various dimensions. Such data contain different types of information in different dimensions and also include time-oriented dimensions and features. Therefore, such kind of data ask for particular interactive visualization methods, which capture all data properties.

We introduce a novel interactive Information Visualization (InfoVis) technique called the Stardinates which aims at visualizing highly structured data, e.g., psychotherapeutic data. The name results from a compound of 'Star' and 'Coordinates' because axes are arranged in a circle. Thus, the visualization looks like a star (compare Figure 1). Each axis represents one attribute of the data so that every data record is visualized by a line (record line) along the corresponding data points. So that a polygon is formed by the lines connecting the data points of one data record. The Stardinates are a hybrid visualization technique combining geometric and glyph-based visualizations. In the following we give a short overview about related InfoVis techniques, explain our background ideas and clarify the concept of the Stardinates based on mathematical definitions in order to outline a foundation for implementation. Thereafter, we sketch interaction techniques, show an application by visualizing psychotherapeutic data derived from a clinical study, and recapitulate the benefits and shortcomings of the Stardinates. Finally, we conclude with an outlook to future work.

2 Related Work

The goal of InfoVis is to promote a more intuitive and deeper level of understanding of the investigational data and to foster new insights into the underlying process [10]. InfoVis is the use of computer-supported, interactive, visual representations of abstract data in order to facilitate cognition. It applies visual processing methods to abstract data and information and is a kind of external cognition — using resources in the world outside the mind in order to expand what the mind can do. An enormous amount of work has been done in the field of InfoVis in the last years. The most prominent ones related to our approach are geometric techniques (e.g., parallel coordinates [7]), Glyphs [6] and similarly icon-based techniques (e.g., Chernoff faces [5]).

The challenges in designing visualization methods for time-oriented psychotherapeutic data is quite specific. The visualization should display the derived qualitative temporal patterns and important quantitative time-oriented values in a task-specific way and provide more information on a user's request. One way to deal with high-dimensional, time-oriented data are multi-views which offer different visualizations of the same data. We developed a multi-view tool, called LinkVis [8] which uses parallel coordinates, Chernoff faces, and scatterplots to provide the user with different views. This multi-views enable the user to prove her or his interpretations by looking for equal patterns in the other visualizations.

Although this multi-view approach is very promising we began to investigate a new visualization technique, since our highly structured data — psychotherapeutic data based on questionnaires — had some specific features the three existing methods used in LinkVis could not handle optimally. Our novel interactive visualization technique called the Stardinates combines features of parallel coordinates and Glyphs and seems more appropriate for this kind of data. We use linking and brushing [1] to enable the user to explore the data efficiently. In contrast to already existing methods, like Star Glyphs [9] applied in the XmdvTool [11], Star Plots [4] and wheel charts, several data records are depicted as lines within one Stardinate and several Stardinates are placed next to each other like Glyphs. In the following we describe this visualization technique in more detail.

3 The Stardinates

3.1 Data Characteristics

Generally, InfoVis deals with the visualization of different types of data from different sources with no inherent spatial meaning. Every visualization is optimized for some types of data. Therefore, we want to give a short description of the data the Stardinates can handle. The Stardinates are suitable for ordinal and nominal values. On the one hand the complexity of the data structure could be very high since every axis of one Stardinate can depict another dimension. On the other hand the amount of data (frequency) and the parameter space is restricted because of the glyph-size of the Stardinates. If the number of lines or values is too high, the shapes of the Stardinates lack of expressiveness so that they are hardly distinguished. We tested the Stardinates with up to 20 axes considered as the maximum of axes of one Stardinate. Therefore, this method is not suitable for high frequency data. This data classification corresponds to the characteristics of the data we derived from a clinical study based on questionnaires, in particular, psychotherapeutic data. The values which represents the answers in the questionnaires usually range from '0' to '6', '14' is the maximum.

3.2 The Method

We use axes (x_1 to x_n) to represent the various dimensions of our data. The axes are arranged in a circle. Each Stardinate consists of a number of axes with scales and labels. The data values are displayed as points on the scales. In order to emphasize the data points, two data points on adjacent axes are connected by a line. So one data record is displayed by a line with its vertices on the axes, we call this record line. This results in the typical star-shape which is perceived efficiently. In contrast to the parallel coordinates we use closed star-shapes. The last axis x_n is placed next to the first axis x_1 and its values are related. The same property holds for the other adjacent axes. According to the Gestalt laws [2] such shapes are perceived intuitively. Like Glpyhs a number of Stardinates are displayed next to each other in order to extract one dimension. Typically meta dimensions like time or other dimensions of special interest are extracted this way to make the visualization more understandable as we will show in the application beneath (compare Figure 1). Generally, the search for multivariate relations in n-dimensional data is transformed into a 2-D pattern recognition problem. Large shapes indicate high values, small shapes imply low values.

The record line is characterized by color, saturation, brightness, sharpness, thickness and mode (dotted, etc.) of line. This is the basis for selecting and filtering data by the user. To provide the user with detailed information every record line offers context information on demand if the user focuses on it. In contrast to the parallel coordinates equal values cannot be recognized by straight lines. In order to ease the recognition we use a reference line which connects equal values on the axes on demand. The user can display, move, or hide the reference line according to her or his needs.

In order to enable the user to explore the data interactively, the visualization tool needs to offer direct interaction techniques based on the users' tasks and combined with immediate respond times.

In the following we specify the Stardinates in detail by mathematical definitions for the axes, the data lines (record lines) and its vertices (data points).

4 Formal Description of the Stardinates

The number of the axes of one Stardinate is derived from the number of dimensions of the visualized data. So we define the axes by:

$$x_1, \ldots, x_n \quad \text{with} \;\; n \in \mathbb{N} \;\; \text{and} \;\; 3 \leq n \leq n_{max} \qquad (1)$$

n_{max}: As mentioned above an upper limit for the number of dimensions to $n_{max} \leq 20$ seems useful.

The similarity the Stardinates and the polar coordinates is obvious. From a geometrical point of view the axes of the Stardinates are placed on the coordinate lines which emanate from the pole of the polar coordinates radially. We start with x_1 at twelve o'clock position and then place the rest of the axes clockwise. So we describe the angle for each axis by:

$$\varphi_{x_k} = \frac{\pi}{2} - (k-1) \cdot \frac{2\,\pi}{n} \quad \forall\, k \in \mathbb{N},\; 1 \leq k \leq n_{max} \quad (2)$$

The axes have a specific length according to the size of the window dedicated to one Stardinate. Based on the size of the window $window_x$(width),$window_y$ (height) we define the length of the axis by:

$$l = \frac{3}{7} \cdot \min(window_x, window_y) \tag{3}$$

The coefficient $\frac{3}{7}$ results from the ideal case $n = 4, 6, 8, \ldots$ when the center (pole) of the Stardinate is placed at $\left(\frac{window_x}{2}; \frac{window_y}{2}\right)$ in the center of the window. If $n = 3, 5, 7, \ldots$ the lower space is used inefficiently because the lower axes are inclined and therefore, do not need the same vertical space like straight axes. But this is nearly irrelevant when $n \geq 7$ because then it converges rapidly to the ideal case. In the center of the window $\frac{1}{7}$ of size of the window is reserved for labels etc. That means we shift each axis out for $\frac{1}{6}$ of the axis length. Together with the $\frac{6}{6}$ of the axis we get the total size of $\frac{14}{6}$. The reciprocal value gives us the length of the axis in relation to the size of the window. Finally, we derive the definition of the axes by:

$$x_k: \ \frac{l}{6} \leq \varrho_{x_k} \leq \frac{7\,l}{6} \quad \text{and } \varphi_{x_k} \text{ from (2)} \tag{4}$$

Next we define the data points and the lines connecting the points on adjacent axes. Therefore, we specify minimum and maximum scale values for each axis: $x_{i\ min}$ = minimum scale value on axis x_i and $x_{i\ max}$ analog.

To display the value of a specific data record (data record a) we need to calculate the position in polar coordinates. So our input $d_{a,i}$ is a value of the dimension i of the data and results in the point $D_{a,i}$ which is displayed on the axis x_i. In our definition we distinguish the negative and the positive sections of the scale in order to enable the user to select different scales or different min/max values for every axis.

We derive the position of the point at the coordinates:

$$D_{a,i}(\varrho_{a,i}; \varphi_{x_i}) = D_{a,i}\left(r_{fix} + \frac{(d_{a,i} - x_{i\ min}) \cdot l}{x_{i\ max} - x_{i\ min}}; \varphi_{x_i}\right) \tag{5}$$

with $r_{fix} = \frac{l}{6}$ analog to (4)

The data points of one record on adjacent axes x_i and x_j are connected by a line. For the values of $D_{a,i}$ on axis x_i and $D_{a,j}$ on axis x_j we define the line $l_{a,i,j}$ connecting these two points:

$$D_{a,i}, D_{a,j} \quad \text{with} \quad 1 \leq i \leq n \tag{6}$$

$$\text{and} \quad j = \begin{cases} i+1 & \text{if } i \leq n-1 \\ 1 & \text{if } i = n \end{cases} \tag{7}$$

$l_{a,i,j}$ given by $\varrho_{a,i,j}$ within the range of $\varphi_{a,i,j}$:

$$\varrho_{a,i,j} = \frac{\varrho_{a,i} \cdot \varrho_{a,j} \cdot \sin(\varphi_{x_j} - \varphi_{x_i})}{\varrho_{a,i} \cdot \sin(\varphi_{a,i,j} - \varphi_{x_i}) - \varrho_{a,j} \cdot \sin(\varphi_{a,i,j} - \varphi_{x_j})} \tag{8}$$

$$\varphi_{a,i,j} = \begin{cases} \varphi_{x_i} \geq \varphi_{a,i,j} \geq \varphi_{x_j} & \text{if } i < j \\ \varphi_{x_i} \geq \varphi_{a,i,j} \geq -\frac{3\pi}{2} & \text{if } i > j \end{cases} \tag{9}$$

In only one case the value of i is higher than the value of j. That happens when comparing the axis x_n with x_1.

Equation (8) describing the data line is undefined if the denominator is '0'. By an indirect proof we can show that this never happens if we stick to the formulas above (hint: $n \geq 3$ and $r_{fix} > 0$). Thus, the definition of the line connecting two data points of the Stardinates is proved to be valid.

This mathematical definition describes how Stardinates are constructed and provides us with a good starting point for the implementation of a Stardinates tool.

5 Interacting with the Stardinates

From a technical point of view we classify users' interaction with the visual representation by the object of interaction. We think this is a good basis for both, the implementation and the concept of the user interface. The user could (a) manipulate the axes, (b) the record line, (c) one Stardinate or (d) all Stardinate at once. Manipulation of one object could cause manipulations of the dependent objects. Examples for (a) manipulating an axis are: Adding or excluding dimensions respectively the corresponding axes; Arranging axes; Changing the orientation of an axis; Defining the scales and minimum and maximum values; Using sliders attached to the axes in order to filter interesting data; (b) Manipulating a record line means: Highlighting or fading out one or more record lines by filtering; Linking and brushing (selecting record lines in one Stardinate causes the highlighting of the corresponding record lines in the other Stardinates automatically); Comparing with the reference line. Every manipulation of an axis results in mutations of record lines. (c) Examples for handling one Stardinate: Magnifying and Zooming out. Finally, (d) the configuration of the Stardinates includes the space per Stardinate (see (3)); The color, thickness, and mode of the record lines (described in (8)) and axes; Saving of interesting bundles of dimensions; Comparing or saving of visualizations.

6 Application

We use the Stardinates to visualize psychotherapeutic data derived from a clinical study on anorectic girls. In

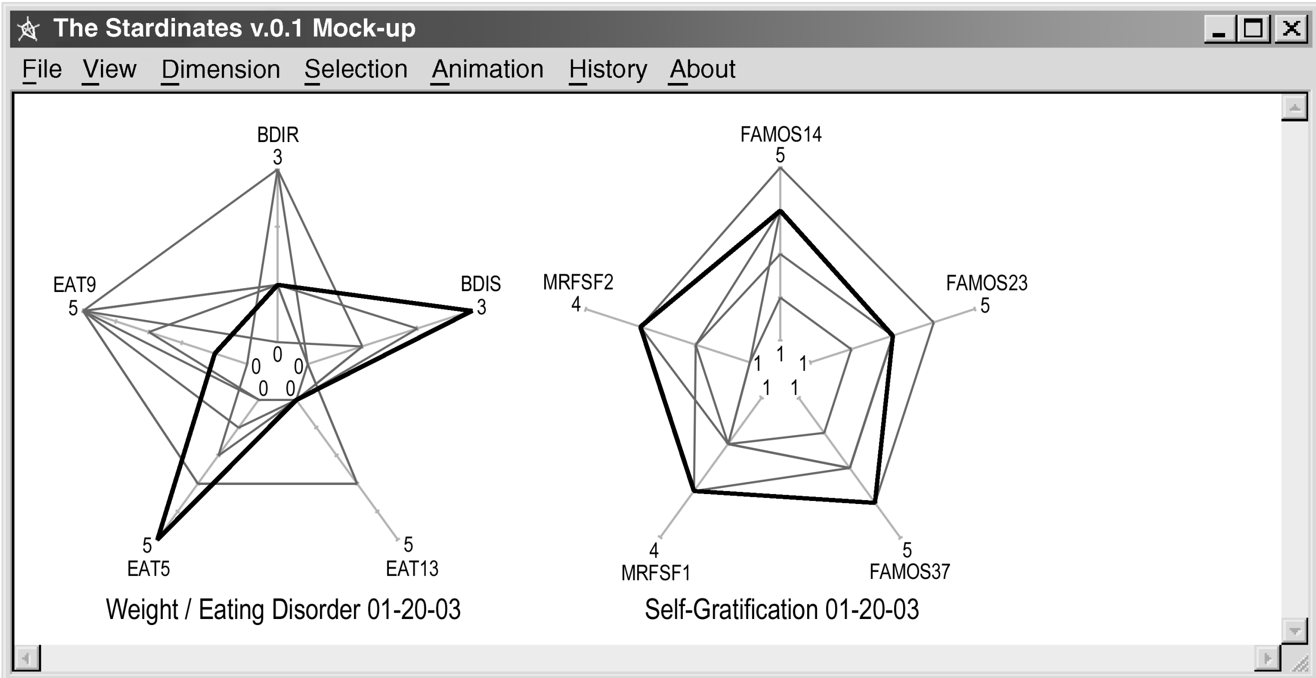

Figure 1. Mock-up of the Stardinates tool.

close cooperation with two psychologists we are developing a visualization tool which aims to support the evaluation and analysis of the therapeutic progress with focus on the differences of group and individual therapy. Above we show a mock-up of this tool. The clinical study works with questionnaires—about 150 per patient. Each questionnaire consists of about 40 questions. Since some questionnaires are used several times, we reduce the number of dimensions respectively. This results in about 1000 dimensions per patient and is described as highly structured data. The questions have unique abbreviations which can be used as labels of the axes. The huge amount of questions makes up a number of dimensions to be visualized. We have defined meaningful bundles of questions which show the state of the patients according to significant characteristics of the ailment and the therapeutic progress, e.g. depression, loss of weight, self-confidence, social context, eating disorder, etc. So far 7 patients have started with individual therapy.

The data derived from these questionnaires mainly ranges from the value '0' to '6' which is a scale from 'No' to 'Yes' and is classified as ordinal data. The granularity of the data varies. Some of the questions have negative polarity, therefore, their axes are flipped in order to fit in one Stardinate together with questions with positive polarity. Analysis and exploration of these data needs targeted interaction which implies knowledge on the underlying questionnaires and their meanings. The Stardinates support the user by intuitive distinguishable shapes, however, in order to really understand the data users need experience in psychology.

During the next two years the therapies of 60 patients will be evaluated. For quantitative analysis applicable techniques based on statistic methods already exist. Since the psychologists plan to pay detailed attention to qualitative exploration and analysis they need new techniques. Visualization offers excellent opportunities for efficient perception and recognition of relevant information because it makes use of our special visual abilities, e.g. pre-attentiveness. In this case visualization has crucial advantages over text or tables.

In the following we show a mock-up (Figure 1) which consists of two Stardinates. Each visualizes the state of seven patients according to five dimensions / questions. The right Stardinate deals with self-gratification abilities of the patients. The other one depicts data of weight / eating disorders. Both show data from Jan. 20th. One record line—patient id. 7 is marked in both Stardinates. This is well-known as linking and brushing. Surely, the Stardinates are not intended as static images. Interacting with them is absolutely necessary in order to build hypotheses and get more insight into the data under investigation. However, this mock-up gives an impression of our technique. Each axis has its own scale and minimum and maximum values drawn from the questionnaires. To normalize the parameter space the user can change this scales adequately.

The Stardinate on 'Self-Gratification' is based on the following questions:

FAMOS14: 'To relax is': (1) totally unimportant (2) unimportant, (3) does not matter (4) rather important to (5) extremely important to me.
FAMOS23: 'To do something just for me is': analog.
FAMOS37: 'To treat oneself to something is': analog.
MRFSF1: 'I treat myself to tranquillity and recreation': (1) no, (2) rather no, (3) rather yes, (4) yes.
MRFSF2: 'I pay close attention to my body': analog.
The axes starting in twelve o' clock represents these questions.

The Stardinate on 'Weight / Eating Disorder' is based on the following questions:
BDIR: 'loss in appetite': (0) no, (1) rather no, (3) rather yes, (4) yes.
BDIS: 'loss in weight': analog.
EAT13: 'I feel sick after eating': (0) never, (1) seldom, (2) sometimes, (3) often, (4) very often, (5) always.
EAT5: 'I avoid to eat when i am hungry': analog.
EAT9: 'I look out for the calories of my food': analog.

The data result in easily perceivable shapes which needs to be interpreted. Since we use five axes here, the data record of each patient is depicted as a pentagon. The Stardinate on self-gratification shows that patient id. 7 has high values, compared to other patients the highest values at FAMOS23, MRFSF1, and MRFSF2. The values of the other axes are also in the upper area. This means that this patients has – especially compared to other patients of this study – in principle the attempt and possibly the ability of self gratification because she rates them with high priority. However, the interpretation of the Stardinate about weight and eating disorders is more complicated. The patients rates BDIS and EAT5 very high, the other axes show the lowest value of '0'. In contrast to other patients, patient id. 7 has low values at EAT9 and BDIR. That means that calories are not important to her and she did not feel the loss of appetite but lost weight. This corresponds with her answer about avoiding food although she feels hungry. These conclusions are a first step into the analysis of the state of this patient. So far we do not have more data, but it would be very interesting to show the changes over time by other Stardinates positioned beneath. Moreover, other meaningful bundles of dimensions offer additional opportunities for informed interpretations.

7 Shortcomings and Benefits

Some problems and shortcomings of the Stardinates are not solved so far. Similarly to the parallel coordinates record lines with equal values cover each other. So the user can not see, if a line represents one ore more data records. Another problem occurs if the data are incomplete. This could happen easily with data derived from questionnaires. A possible approach is to add an extra point on the scale for missing values. However, this does not seem to be the best solution.

To mention two other restrictions: The parameter space and the number of axes is limited. The number of Stardinates itself is also limited, because it is not possible to compare hundreds of Stardinates efficiently.

However, the Stardinates offer also many benefits. Visualizations have several advantages [3] from a cognitive point of view. Stardinates seem to support some of these cognitive mechanisms in a fairly efficient manner. They represent some Gestalt principles very well, especially the principles of Closure and 'Prägnanz'. As a consequence, Stardinates form very distinct and memorable patterns which make abstraction and aggregation much easier. They also support perceptual inferences because the meaning of the elements of Stardinates are intuitively fairly clear. In contrast to that, Glyphs, for example, require much more effort to learn the meaning of the elements of visualization. The possibility to manipulate the data and interact with the form of representation forms another considerable advantage of Stardinates. Manipulating representations enables users to produce visualizations which are specifically tailored to their needs. In many cases, such representations are more meaningful for the users and better adapted to help users to solve their problems. The interactivity of Stardinates also helps users to formulate and test hypotheses, thus gaining deeper insights into the subject at hand. Stardinates share some of these advantages with other methods of visualization. What is perhaps unique about Stardinates is the combination of these advantages.

8 Conclusion and Future Work

Collecting data is a time consuming task and asks for sophisticated exploration techniques to gain more information and knowledge. In particular time-oriented and highly structured data, which contain of a complex data structure, but only a few instances, need specific exploration techniques. We have presented an interactive InfoVis method, called Stardinates, which addresses the particular features of such highly structured data.

The usefulness of the Stardinates is illustrated with psychotherapeutic data collected in a clinical study on anorectic girls. That study provides complex and time-dependent data, which are difficult to explore by descriptive and classical statistical methods.

As demonstrated in Figure 1, the study of Stardinates showed promising results and the cooperating psychologists see various application areas for this InfoVis technique within their clinical study of anorectic girls. Therefore, in the next step we plan detailed user testing. We will do a study with 60 subjects to compare the effectiveness of the Stardinates with other visualization techniques. Additionally, we will perform ethnographic studies to collect more information about the real users' needs. The findings will influence and will be assembled in the development phase of the Stardinates tool. Within cyclical evaluation and implementation phases we will develop a visualization tool, which will support the whole psychotherapeutic processes as well as facilitate to analyze the effectiveness, benefits, and limitations of various treatment options and to compare the treatment courses.

From the technical point of view, we will investigate in more advanced techniques to capture the temporal properties. One possible extension is to integrate animated Stardinates. Such an animated view will show the changes of the record line over time for one patient. According to the Gestalt laws the movements of record line can be perceived intuitively.

Finally, we are planning to evaluate the Stardinates with traditional InfoVis methods, which address similar features. These methods will be parallel coordinates, Chernoff faces and their derivatives thereof. We are hopeful that we can examine and show, that the interactive Stardinates are more intuitive, easier to understand, and support faster capturing of coherences.

9 Acknowledgements

The authors would like to thank the psychologists Susanne Ohmann and Christian Popow. Furthermore many thanks to Markus Rester for interesting discussions about our visualization technique and valuable LaTeX support.

References

[1] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Advanced Visual Interfaces (AVI2000)*, pages 110–119. ACM Press, 2000.

[2] J. G. Benjafield. *Cognition*. Prentice-Hall International Editions, 1992.

[3] S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization*, chapter 1, pages 1–34. Morgan Kaufman, 1999.

[4] J. Chambers, Cleveland, K. W., and P. B., Tukey. Graphical methods for data analysis. Wadworth.

[5] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68:361–368, 1973.

[6] M. C. Chuah and S. G. Eick. Glyphs for software visualization. In *5th International Workshop on Program Comprehension (IWPC '97) Proceedings*, pages 183–191. IEEE Computer Society Press, Dearborn, Michigan, May 1997.

[7] A. Inselberg. The plane with parallel coordinates. *Special Issue on Computational Geometry*, 1:69–97, 1997.

[8] M. Lanzenberger, S. Miksch, S. Ohmann, and C. Popow. Applying information visualization techniques to capture and explore the course of cognitive behavioral therapy. In *Proc. of the ACM Symposium on Applied Computing, SAC 2003*, pages 268–274, 2003.

[9] J. Siegel, E. Farrell, R. Goldwyn, and H. Friedman. The surgical implication of physiologic patterns in myocaridal infarction shock. *Surgery*, 72:126–141, 1972.

[10] E. R. Tufte. *Visual Explanations*. Graphics Press, Cheshire, CT, 1997.

[11] M. O. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of Visualization 1994*, pages 326–333, 1994.